

**XIII CONGRESSO**

**NACIONAL DE GEOTECNIA**

Pensar e Construir com a Natureza. Uma Visão para a Engenharia

**VI CONGRESSO**

**LUSO-BRASILEIRO DE GEOTECNIA**

Risco e Segurança

**LISBOA 2012**

*Título*

XIII CONGRESSO NACIONAL DE GEOTECNIA  
Pensar e Construir com a Natureza  
Uma Visão para a Engenharia

VI CONGRESSO LUSO-BRASILEIRO DE GEOTECNIA  
Risco e Segurança

*Autores*

Vários

*Pré-Impressão | Impressão | Acabamento*  
Cor Comum - Serviços Gráficos, Lda.

*Data*

Abril de 2012

*ISBN*

978-972-98781-8-3

*Depósito Legal*

342944/12

# Utilização de Máquina de Vetores de Suporte na Previsão do Comportamento Mecânico de Formulações Laboratoriais de Solo-Cimento para Colunas de Jet Grouting

## Support Vector Machines on Mechanical Behaviour Prediction of Soil-Cement Laboratory Formulations to Jet Grouting Columns

Tinoco, Joaquim; *Universidade do Minho, Guimarães, Portugal*, [jabtinoco@civil.uminho.pt](mailto:jabtinoco@civil.uminho.pt)  
Gomes Correia, António; *Universidade do Minho, Guimarães, Portugal*, [agc@civil.uminho.pt](mailto:agc@civil.uminho.pt)  
Cortez, Paulo; *Universidade do Minho, Guimarães, Portugal*, [pcortez@dsi.uminho.pt](mailto:pcortez@dsi.uminho.pt)

### 1 - INTRODUÇÃO

No contexto da tecnologia de Jet Grouting (JG) (Shibazaki, 2003) a principal desvantagem está relacionada com a ausência de modelos de dimensionamento racionais para a previsão das propriedades mecânicas das misturas de JG. Atualmente, esta tarefa é essencialmente realizada considerando os resultados de ensaios laboratoriais e modelos empíricos. No entanto, nas fases iniciais de projeto, principalmente em obras geotécnicas de pequena envergadura, a escolha dos valores a adotar para cada um dos parâmetros é baseada em informação muito escassa. Assim, o uso de dados de projetos já realizados pode ser visto como uma alternativa para ultrapassar este problema. Por conseguinte, e tendo presente a elevado potencial e versatilidade da tecnologia de JG bem como o seu papel em importantes obras geotécnicas, torna-se evidente a necessidade de desenvolver métodos racionais e fiáveis para a previsão das propriedades físicas e mecânicas do material JG. Nesse sentido, a aplicação de técnicas de data mining (DM) (Fayyad et al., 1996) a dados bem organizados coletados em grandes obras geotécnicas pode fornecer uma importante base para o desenvolvimento de modelos que poderão ser muito úteis em futuros projetos. O presente trabalho atualiza os modelos apresentados em Tinoco et al. (2010) para a previsão da resistência à compressão uniaxial ( $Q_u$ ) e deformabilidade de formulações laboratoriais de jet grouting (FLJG) ao longo do tempo através da aplicação de técnicas de DM, mais precisamente através do uso de máquinas de vetores de suporte (MVS). Assim, foi treinada uma MVS para prever a resistência à compressão uniaxial ( $Q_u$ ) de FLJG, propriedade correntemente utilizada para medir a capacidade resistente das colunas de fundação. Nas estruturas geotécnicas onde os estados limites de utilização são requeridos, as propriedades de deformabilidade do solo tratado precisam também de ser quantificadas. Para este efeito, foram também treinadas MVSs no sentido de prever os diferentes módulos de deformabilidade, isto é, módulo de elasticidade inicial ( $E_0$ ), módulo de deformabilidade tangente a 50% da tensão de rotura ( $E_{tg50\%}$ ), módulo de deformabilidade secante a 50% da tensão de rotura ( $E_{sec50\%}$ ) e módulo de deformabilidade máximo ( $E_{max}$ ) de FLJG ao longo do tempo. Foi também realizada uma análise de sensibilidade detalhada (Cortez e Embrechts, 2011) no sentido de identificar as variáveis chave e qual o seu efeito no estudo das propriedades mecânicas de FLJG, contribuindo para uma melhor interpretação dos modelos.

### 2 - RESULTADOS E CONCLUSÕES

Os resultados obtidos mostraram que o algoritmo MVS é capaz de aprender as complexas relações entre as propriedades mecânicas de FLJG e as diferentes variáveis intervenientes. Treinando este algoritmo obtiveram-se modelos capazes de estimar com elevado grau de precisão ( $R^2$  próximos da unidade) a  $Q_u$  e  $E_0$  entre os 3 a 56 dias, e o  $E_{sec50\%}$ ,  $E_{tg50\%}$  e  $E_{max}$  entre 28 e 84 dias de FLJG preparadas com solos argilosos contendo diferentes percentagens de areia, silte, argila e matéria orgânica. O elevado desempenho dos modelos é verificado pela semelhança entre os valores previstos e os experimentais. A construção das curvas REC (*Regression Curve Characteristic*) para os diferentes modelos permitiu ainda observar um rápido aumento da precisão dos modelos em função do erro tolerado. Após realização de uma análise de sensibilidade sobre os diferentes modelos treinados e quantificada a importância relativa de cada atributo (ver Figura 1) foi possível verificar que a relação entre a porosidade da mistura e o teor volumétrico de cimento ( $n/(C_w)^d$ ) é a variável chave no estudo das propriedades mecânicas de FLJG. No caso do estudo da  $Q_u$ , a variável tempo ( $t$ , dias) e o teor em cimento da mistura ( $C$ ) são também variáveis muito relevantes. A razão água/cimento ( $A/C$ ), apesar de identificada como a quarta variável com maior impacto na previsão da  $Q_u$  de FLJG, deverá também ser levada em consideração no estudo do comportamento mecânico deste material pois apresenta uma forte interação com a variável  $t$  (14%). Verificou-se inclusive que misturas com elevada  $A/C$  tendem a estabilizar para idades mais recentes. Quando o objetivo passa por estimar a deformabilidade de FLJG, as propriedades do solo natural, principalmente o seu teor em argila, seguem a relação  $n/(C_w)^d$  nas variáveis com maior impacto. Uma observação também interessante, no caso do estudo da deformabilidade, prende-se com o facto de a idade da mistura apresentar uma importância relativamente pequena, exceto no estudo do  $E_0$ . Esta constatação pode ser justificada pelo facto de esta variável apresentar uma forte interação com a relação



$n/(C_{IV})^d$ . Da análise da Figura 1 importa ainda a referir o reduzido impacto das propriedades do solo no estudo da resistência quando comparado com o estudo da deformabilidade. A construção das curvas VEC (*Variable Effect Characteristic*) para as variáveis chave no estudo da  $Q_u$ , evidenciou um efeito predominantemente não linear. No caso da variável tempo, o seu efeito segue uma lei exponencial, corroborando assim o conhecimento empírico relativo a misturas solo-cimento. Isto é, que  $t$  tem um efeito mais preponderante nas idades iniciais, tendendo posteriormente a estabilizar.

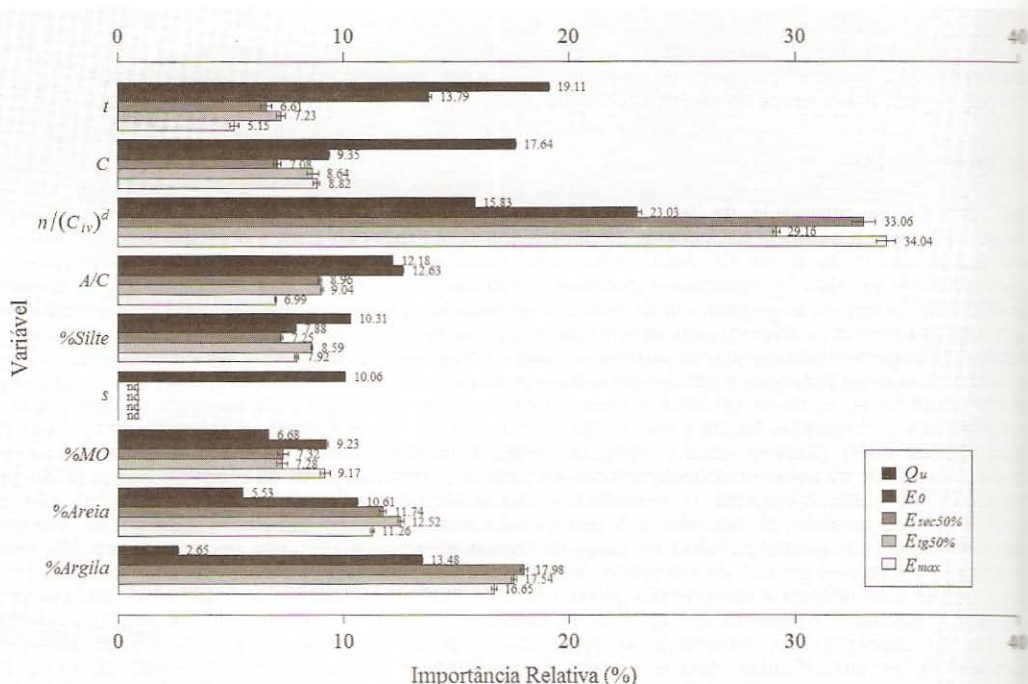


Figura 1 – Importância Relativa de Cada uma das Variáveis no Estudo das Propriedades Mecânicas de FLJG

Face ao excelente desempenho dos modelos treinados, não apenas em termos de capacidade previsional mas também em termos de coerência com o atual conhecimento empírico relativo a FLJG, pode-se concluir que tais modelos podem ser utilizados para descrever com elevado grau de fiabilidade o conhecimento empírico relativo a este tipo de material e servir de apoio ao dimensionamento das respetivas propriedades mecânicas (resistência e deformabilidade). Julga-se que o estudo realizado é uma contribuição para o aumento do conhecimento sobre misturas de solo-cimento para obras de JG, resultando numa economia de formulações laboratoriais ao nível do projeto.

Em trabalhos futuros, estas ferramentas serão aplicadas no desenvolvimento de modelos racionais de previsão das propriedades mecânicas (resistência e deformabilidade) de amostras extraídas diretamente de colunas de JG e ainda na previsão do diâmetro das respetivas colunas.

## REFERÊNCIAS

- Cortez, P. e Embrechts, M. (2011). Opening Black Box Data Mining Models Using Sensitivity Analysis. *Proceedings of the 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2011)*. IEEE, Paris, França, pp.341-348.
- Fayyad, U., Piatetsky-Shapiro, G. e Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*. ACM, Vol.39(11), pp. 27-34.
- Shibasaki, M. (2003). State of practice of jet grouting. *Proceedings of Grouting and Ground Treatment: 3rd International Specialty Conference on Grouting and Ground Treatment*. ASCE, New Orleans, Louisiana, USA, pp.198-217.
- Tinoco, J., Gomes Correia, A. e Cortez, P. (2010). Aplicação de Técnicas de Data Mining na Previsão do Comportamento Mecânico de Colunas de Jet Grouting. *Atas do XII Congresso Nacional de Geotecnia*. A. Gomes Correia et al. (Eds.). Guimarães, Portugal, pp. 2167-2176.

# Utilização de Máquina de Vetores de Suporte na Previsão do Comportamento Mecânico de Formulações Laboratoriais de Solo-Cimento para Colunas de Jet Grouting

## Support Vector Machines on Mechanical Behavior Prediction of Soil-Cement Laboratory Formulations to Jet Grouting Columns

Tinoco, Joaquim; *Universidade do Minho, Guimarães, Portugal*, [jabtinoco@civil.uminho.pt](mailto:jabtinoco@civil.uminho.pt)  
Gomes Correia, António; *Universidade do Minho, Guimarães, Portugal*, [agc@civil.uminho.pt](mailto:agc@civil.uminho.pt)  
Cortez, Paulo; *Universidade do Minho, Guimarães, Portugal*, [pcortez@dsi.uminho.pt](mailto:pcortez@dsi.uminho.pt)

### RESUMO

Caracterizada pela sua grande versatilidade, a tecnologia de jet grouting (JG) é atualmente um dos métodos de tratamento de solos moles mais utilizados, nomeadamente em importantes obras geotécnicas. Contudo, o dimensionamento das propriedades físicas e mecânicas do novo material resultante do tratamento é essencialmente suportado por abordagens empíricas frequentemente muito conservativas. Como consequência, a eficiência técnica e económica do tratamento poderá ser comprometida. No sentido de colmatar esta lacuna, no presente artigo é apresentada uma abordagem racional para a previsão da resistência e deformabilidade de formulações laboratoriais de JG através da aplicação de técnicas de *data mining*, mais concretamente através do uso de máquinas de vetores de suporte. É ainda realizada uma análise de sensibilidade detalhada, tendo por objetivo identificar as variáveis chave e qual o seu efeito no estudo das propriedades mecânicas do material em estudo.

### ABSTRACT

Characterized by its great versatility, jet grouting (JG) technology is one of the most soft soil improvement methods currently applied, namely in important geotechnical works. However, physical and mechanical properties design of the new obtained material is essentially supported by empirical approaches often too conservative. As a result, the efficiency of the treatment can be technically and economically compromised. In order to overcome such drawback, in the present paper, a rational approach to predict strength and stiffness of jet grouting laboratory formulations by application of data mining techniques is proposed, particularly by means of support vector machine algorithm. It is also performed a detailed sensitivity analysis in order to identify the key variables and its effect in the study of these mechanical properties.

### 1 - INTRODUÇÃO

Para servir como o solo de fundação, o solo tem de apresentar as propriedades geotécnicas adequadas. No entanto, devido à crescente urbanização e industrialização, a disponibilidade de solos que cumpram tais requisitos é cada vez mais escassa. Como consequência, cada vez mais há a necessidade de construir sobre solos moles caracterizados por elevada plasticidade, percentagem de finos e índice de vazios, baixa resistência e elevada compressibilidade (Liu et al., 2008). No sentido de dotar os solos com as propriedades adequadas, um dos métodos mais populares consiste em proceder à sua mistura com uma calda de cimento que é injetada no interior do solo natural. Hoje em dia existem diversos métodos de melhoramento de solos moles (Nikbakhtan et al., 2010b), destacando-se a tecnologia de *Jet Grouting* (JG) pela sua forte versatilidade. Esta tecnologia permite melhorar as propriedades físicas e mecânicas de diversos tipos de solo (desde solos finos até solos granulares), obter diferentes formas de tratamento (colunas, painéis, etc.). Além disso, a quantidade de equipamentos/maquinaria necessários para a sua aplicação é reduzida e quando comparada com métodos semelhantes é economicamente atrativa. Esta tecnologia foi introduzida pela primeira vez pelos irmãos Yamakoda na década de 60 e desde então tem sido largamente utilizada em diversas partes do mundo (Xanthakos et al., 1994).

Conceptualmente a tecnologia de JG tem por objetivo produzir uma mistura solo-cimento, também designada por *soilcrete*, com melhores propriedades físicas e mecânicas (resistência e deformabilidade) em relação ao solo natural. Em termos de processo construtivo, o melhoramento é efetuado através da injeção de calda de cimento, eventualmente em conjunto com outros fluidos (ar e/ou água) a elevadas pressões e velocidade no interior do solo natural. Os fluidos são injetados através de pequenos orifícios colocados numa haste que, após introduzida até à profundidade pretendida é animada de movimento rotacional e lentamente removida até à superfície. Em função do número de fluidos injetados, existem três tipos de sistemas correntemente aplicados (Xanthakos et al., 1994). No sistema de *jet simples* é apenas injetado calda de cimento a elevadas pressões e velocidades no interior do solo; o sistema de *jet duplo* é em tudo semelhante ao sistema simples, diferindo apenas no facto de o jacto de calda ser

envolvido por ar comprimido o que permite aumentar a área de tratamento; finalmente no sistema de *jet triplo* a erosão do solo é efetuada por um jacto de água a elevada pressão envolvido por ar, sendo a mistura realizada por um jacto de calda de baixa pressão. Além destes três sistemas principais, existe também o sistema *Xjet*, também conhecido por *cross-jet*. Segundo este sistema um par de jatos de água envolvidos por ar intersectam-se permitindo obter um diâmetro nominal de 2 metros em qualquer tipo de solo (Shibazaki, 2003), sendo a mistura realizada por um jacto de calda de baixa pressão.

Em resultado da grande versatilidade da tecnologia de JG a mesma pode ser utilizada para diversas finalidades, nomeadamente no melhoramento do solo de fundação, estabilização de taludes ou reforço de fundações (Welsh e Burke, 1991). Além disso, tem sido também utilizada na construção de blocos de impermeabilização de solos contaminados bem como na formação de barreiras para o controle da migração de contaminantes na área ambiental (Gazaway e Jasperse, 1992).

Apesar de largamente utilizada, nomeadamente em importantes obras geotécnicas, o dimensionamento da tecnologia de JG carece de abordagens racionais de previsão das propriedades físicas e mecânicas do novo material resultante do tratamento. De facto, o dimensionamento do JG, tanto em obras de grande envergadura como de pequena dimensão onde a informação é escassa, é essencialmente realizado através de métodos empíricos (Croce e Flora, 2001; Kirsh e Wolfgang, 2001) e suportado pelo conhecimento das respetivas empresas. A Figura 1 esquematiza o procedimento de controlo de qualidade normalmente adotado. As principais razões para tal cenário prende-se por um lado, pelo elevado número de variáveis envolvidas as quais apresentam relações complexas ou mesmo de não linearidade entre si e por outro, pela heterogeneidade do solo natural. Como resultado, e uma vez que tais métodos empíricos são frequentemente muito conservativos e com aplicabilidade limitada, a qualidade e a economia do tratamento podem ser comprometidos. Assim, e tendo presente a elevado potencial e versatilidade da tecnologia de JG bem como o seu importante papel em importantes obras geotécnicas, torna-se evidente a necessidade de desenvolver métodos racionais e fiáveis para a previsão das propriedades físicas e mecânicas do material JG.

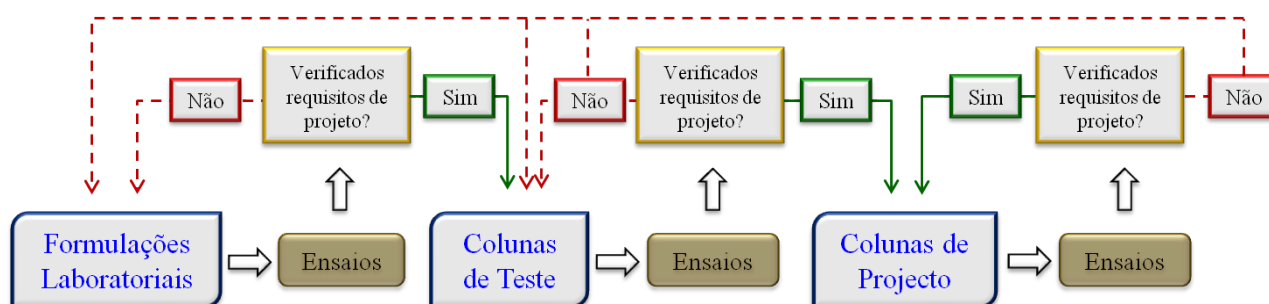


Figura 1 – Processo Empírico de Dimensionamento da Tecnologia de JG

Algumas das expressões matemáticas existentes na bibliografia para previsão das propriedades mecânicas do material JG (Lee et al., 2005; Narendra et al., 1996; Nikbakhtan e Ahangari, 2010a), resultam da aplicação de análises estatísticas tradicionais. Contudo, estas ferramentas não são capazes de dar a resposta adequada perante dados de JG caracterizados por uma elevada complexidade. Uma das melhores alternativas para lidar com um elevado número de variáveis com relações complexas é utilizando um processo automático através da aplicação de ferramentas de inteligência artificial que analisa as linhas da base de dados e extrai padrões e/ou regras para os mesmos. O processo de descoberta de conhecimento em base de dados (DCBD) (ver Figura 2) é um processo iterativo que consiste em várias etapas, onde o *data mining* representa apenas uma dessas etapas (apesar de muito importante) e tem por objetivo extrair padrões/tendências dos dados. De acordo com a metodologia CRISP-DM as seguintes etapas devem ser seguidas (Chapman et al., 2000): 1 – conhecimento do problema; 2 – conhecimento dos dados; 3 – preparação dos dados; 4 – modelação (ou seja, *data mining*); 5 – avaliação; e 6 – desenvolvimento.



Figura 2 – Etapas do Processo de Descoberta de Conhecimento em Base de Dados (adaptado de Fayyad et al., 1996)

Em problemas de regressão e classificação, os quais representam duas importantes tarefas de *data mining* (DM) e requerem uma aprendizagem supervisionada, o modelo é ajustado à base de dados de exemplos através de um mapeamento de  $l$  variáveis de entrada (variáveis independentes) na variável de saída (variável dependente). A construção do modelo pode ser realizada por diferentes algoritmos de DM, cada um com as suas vantagens e limitações. Alguns dos modelos de DM mais interessantes são as Redes Neurais Artificiais (RNAs) e as Máquinas de Vetores de Suporte (MVSs), caracterizadas pela sua flexibilidade e capacidade de aprendizagem não linear. As ferramentas de DM têm sido aplicadas com sucesso em diversas áreas do conhecimento, nomeadamente na área de Geotecnia. Miranda et al. (2011) propôs novos modelos de regressão alternativos usando RNAs para o cálculo analítico dos parâmetros da resistência e deformabilidade de maciços rochosos. Goh e Goh (2007) usou MVSs na avaliação da liquefação sísmica de solos.

O presente trabalho atualiza os modelos apresentados em Tinoco et al. (2010) para a previsão da resistência à compressão uniaxial ( $Q_u$ ) e módulo de deformabilidade tangente a 50% da tensão de rotura ( $E_{tg50\%}$ ) de formulações laboratoriais de jet grouting (FLJG). Esta análise complementar é realizada através da aplicação de uma análise de sensibilidade detalhada sobre um dos modelos de DM, particularmente o algoritmo MVS, visando explorar as variáveis chave e sua contribuição no estudo destas propriedades. Adicionalmente, é apresentada uma abordagem racional, por aplicação do algoritmo MVS, para a previsão dos diferentes módulos de deformabilidade correntemente utilizados, isto é, módulo de elasticidade inicial ( $E_0$ ), módulo de deformabilidade secante a 50% da tensão de rotura ( $E_{se50\%}$ ) e módulo de deformabilidade máximo ( $E_{max}$ ) de FLJG ao longo do tempo. Também aqui é aplicada uma análise de sensibilidade detalhada no sentido de descrever os modelos treinados.

## 2 - MÁQUINA DE VETORES DE SUPORTE

Fundamentado na Teoria da Aprendizagem Estatística, as MVSs, do inglês *Support Vector Machine* foram desenvolvidas por Vapnik (2000), tendo por objetivo resolver problemas de classificação de padrões. Este algoritmo de DM tem mostrado uma elevada capacidade de aprendizagem, mesmo quando usado sobre dados de elevada complexidade (Tinoco et al., 2011a). As MVSs (Chen e Councill, 2003) são uma classe específica de algoritmos caracterizada pelo uso de kernels, ausência de mínimos locais durante a fase de aprendizagem e capacidade de controlo conseguida através da definição da margem e do número de vetores de suporte. Quando comparado com outros tipos de algoritmos, nomeadamente as RNAs, as MVSs representam uma melhoria significativa em termos de funcionalidade. Esta supremacia deve-se ao uso de funções kernel não lineares que implicitamente fazem um mapeamento das entradas (*inputs*) num espaço imaginário de elevada dimensão. Neste espaço imaginário é possível realizar operações lineares que procuram encontrar o melhor plano linear de separação ( $y_i = \omega_0 + \sum_{i=1}^m \omega_i \phi_i(x)$ ), em função do conjunto de vetores de suporte, automaticamente definidos pelo algoritmo e que diz respeito aos exemplos mais relevantes para o conceito a aprender pelo modelo. Assim, apesar de MVSs serem máquinas de aprendizagem linear no espaço imaginário, no espaço real de entrada elas são efetivamente não lineares (ver Figura 3). Todas estas características, juntamente com os promissores resultados empíricos, tem contribuído para a sua crescente popularidade.

As MVSs foram inicialmente propostas por Vladimir Vapnik e seus colaboradores (Cortes e Vapnik, 1995) para resolver problemas de classificação. Mais tarde, após a introdução de uma função perda alternativa proposta por Vapnik (Smola et. al, 1996), chamada função perda  $\varepsilon$ -insensitiva, foi também possível aplicar as MVSs a problemas de regressão (Smola e Schölkopf, 2004).

O desempenho e a capacidade de generalização de uma MVS são fortemente condicionados pelos valores adotados para os hiperparâmetros  $C$  (parâmetro de penalização),  $\varepsilon$  (largura da zona  $\varepsilon$ -insensitiva) e os parâmetros do kernel. O problema da escolha dos valores ótimos para estes hiperparâmetros é ainda agravado pelo facto de a complexidade do modelo MVS (e por conseguinte da sua capacidade de generalização) estar dependente da combinação dos três parâmetros em simultâneo. O parâmetro  $C$  controla a relação entre a complexidade da máquina e o número de pontos não separáveis, podendo ser visto como um parâmetro de regularização (Goh e Goh, 2007). O parâmetro  $\varepsilon$  controla a largura da zona  $\varepsilon$ -insensitiva usada para ajustar os dados de treino. O valor de  $\varepsilon$  pode afetar o número de vetores de suporte usados na construção da função de regressão. Por conseguinte, ambos os valores de  $C$  e  $\varepsilon$  afetam a complexidade do modelo, ainda que de maneira diferente.

A escolha de um tipo de kernel e os parâmetros da respetiva função está normalmente relacionado com o domínio de aplicação e deverá refletir a distribuição dos valores de entrada ( $x$ ) dos dados de treino (Cherkassky e Ma, 2004). No presente trabalho, foi adotado o kernel Gaussiano, uma vez que depende de menos parâmetros quando comparado com outros kernels (e.g. polinomial):

$$K(x, x') = \exp(-\gamma \cdot \|x - x'\|^2), \gamma > 0 \quad [1]$$

Para a definição dos valores dos hiperparâmetros, diversas abordagens tem sido propostas (Huang et al., 2007; Momma and Bennett, 2011; Frohlich and Zell, 2005) no sentido de encontrar o melhor conjunto de parâmetros com o menor esforço, quer em termos de tempo quer em termos computacionais. No presente trabalho, foram adotadas as heurísticas propostas por Cherkassky e Ma (2004). Assim, para  $C$  foi adotado o valor de  $C=3$  e a largura da zona  $\varepsilon$ -insensitiva foi definida de acordo com  $\varepsilon = \hat{\sigma}/\sqrt{N}$ , onde  $\hat{\sigma} = 1.5/N \times \sum_{i=1}^N (y_i - \hat{y}_i)^2$ ,  $\hat{y}_i$  é o valor previsto pelo algoritmo dos 3-vizinhos próximos e  $N$  o número de exemplos. O parâmetro kernel  $\gamma$  foi definido usando uma grelha de procura entre  $\{2^{-15}, 2^{-13}, \dots, 2^3\}$ , e aplicando uma validação cruzada interna (i.e. aplicada sobre os dados de treino) de 10-fold (Hastie et al., 2009).

Todas as experiências foram conduzidas no ambiente estatístico R (R Development Core Team), com o auxílio da biblioteca rminer (Cortez, 2010), a qual é particularmente adequada para o treino do algoritmo MVS e facilita a implementação das abordagens adotadas para a definição dos hiperparâmetros, através do uso de funções avançadas (particularmente as funções "mining" e "fit"). Os dados utilizados para ajustar os modelos (variáveis de entrada) foram previamente normalizadas para uma média de zero e um desvio padrão de um. Posteriormente, antes de analisar as previsões, foi aplicada a respetiva função inversa (Hastie et al., 2009).

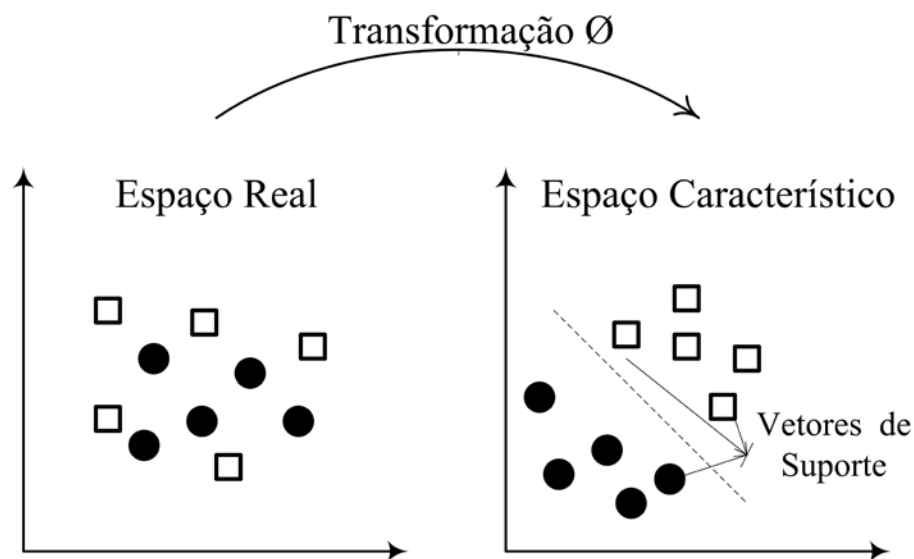


Figura 3 – Exemplo da Transformação MVS (adaptado de Cortez, 2010)

### 3 - AVALIAÇÃO E INTERPRETAÇÃO DOS MODELOS

Em problemas de regressão o principal objetivo passa por definir um modelo que minimize o erro entre os valores experimentais e os valores previstos considerando  $N$  exemplos. Com este pressuposto, no presente trabalho o desempenho dos modelos foi avaliado através do cálculo de três métricas: Desvio Médio Absoluto (MAD), Raiz Quadrada do Erro Médio Quadrado (RMSE) e Coeficiente de Correlação ( $R^2$ ). As expressões matemáticas para cada uma destas métricas são definidas da seguinte forma:

$$MAD = \frac{\sum_{i=1}^N |y - \hat{y}|}{N}; \quad RMSE = \sqrt{\frac{\sum_{i=1}^N (y - \hat{y})^2}{N}}; \quad R^2 = \left( \frac{\sum_{i=1}^N (y - \bar{y}) \times (\hat{y} - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y - \bar{y})^2 \times \sum_{i=1}^N (\hat{y} - \bar{\hat{y}})^2}} \right)^2 \quad [2]$$

Quando as métricas MAD e RMSE apresentam valores reduzidos e o  $R^2$  está perto da unidade é sinónimo de elevado desempenho do modelo. As métricas MAD e RMSE diferem essencialmente pelo facto de esta última ser mais sensível à presença de valores extremos. No sentido de comparar o desempenho de diferentes modelos foram também construídas as respetivas curvas REC (*Regression Curve Characteristic*) para cada um dos modelos treinados. Esta curva representa a percentagem de pontos previstos para uma determinada tolerância (eixo dos yy) em função do erro tolerado (eixo dos xx) (Bi e Bennett, 2003).

A avaliação da capacidade de generalização dos modelos treinados foi efetuada através da implementação da abordagem *Leave-One-Out* (Hastie et al., 2009) e da repetição do processo de treinamento dos modelos 20 vezes. Segundo a abordagem *Leave-One-Out*, cada exemplo é utilizado isoladamente para testar o modelo, sendo os restantes utilizados na fase de treino. No final todos os exemplos são usados tanto para treinar como para testar o modelo. A principal desvantagem desta abordagem prende-se com a grande necessidade de recursos computacionais, uma vez que  $N$  modelos



têm de ser ajustados. A capacidade de generalização final é avaliada através do cálculo das métricas MAD, RMSE e  $R^2$  para todos os  $N$  exemplos de teste.

Ao avaliar o desempenho de um modelo, além da sua capacidade de previsão, é também extremamente importante a sua interpretabilidade. Este aspeto é apontado frequentemente como a principal desvantagem de modelos definidos por aplicação de técnicas de DM. Isto porque as expressões matemáticas que suportam estes modelos são de elevada complexidade e como tal muito difíceis de compreender. Assim, torna-se necessário extrair informação destes modelos num formato mais acessível. Cortez e Embrechts (2011) propõem a aplicação de análises de sensibilidade detalhadas no sentido de extrair informação compreensível dos modelos. Uma análise de sensibilidade é um procedimento que é aplicado após a fase de treino e analisa a resposta do modelo quando uma determinada variável de entrada é alterada. Este procedimento pode ser aplicado a qualquer modelo de DM e permite quantificar a importância relativa de cada uma das variáveis de entrada do modelo bem como medir o efeito médio de determinado atributo na respetiva variável de saída. Esta quantificação é determinada através de sucessivamente fixar todas as variáveis num determinado valor (e.g. o respetivo valor médio), exceto um atributo de entrada que toma todos os valores do seu domínio ( $x_a \in \{x_1, \dots, x_L\}$ ), com ( $j \in \{1, \dots, L\}$ ) níveis. A resposta obtida ( $\hat{y}_{a,j}$ ) é armazenada. Elevadas variações na resposta indicam uma elevada importância da variável em estudo. Em particular, e seguindo os resultados obtidos por Cortez e Embrechts (2011), foi adotado a medida gradiente ( $S_a$ ) para medir a relevância ( $R_a$ ) do atributo  $x_a$  (quanto maior o gradiente maior a importância do atributo):

$$R_a = S_a / \sum_{i=1}^L S_i \times 100(\%), \text{ onde } S_a = \sum_{j=2}^L |\hat{y}_{a,j} - \hat{y}_{a,j-1}| / (L - 1) \quad [3]$$

Para uma análise mais detalhada da influência das variáveis de entrada foram construídas as curvas VEC (*Variable Effect Characteristic*) propostas por Cortez e Embrechts (2011). Para uma determinada variável de entrada, a curva VEC representa o valor do atributo para o nível  $L$  (eixo dos  $xx$ ) versus a resposta da análise de sensibilidade (eixo dos  $yy$ ). No presente trabalho foi adotado  $L=12$ . Além desta análise, e tendo como objetivo realizar uma interpretação mais realista do modelo, foi aplicada uma análise de sensibilidade bidimensional. Aqui são alteradas em simultâneo duas variáveis e medida a respetiva resposta. Com os valores guardados é possível construir superfícies VEC (Cortez e Embrechts, 2011).

#### 4 - BASE DE DADOS DE FORMULAÇÕES LABORATORIAIS

Os dados utilizados durante a fase de treino/teste dos modelos resultaram de um programa experimental laboratorial, realizado na Universidade do Minho, que tinha por objetivo analisar a influência de diversas variáveis no comportamento mecânico de formulações laboratoriais de JG. No presente trabalho estuda-se a resistência à compressão uniaxial, bem como os diferentes módulos de deformabilidade de FLJG (ver Figura 4) que podem ser determinados num ensaio de compressão não confinada (Correia et al., 2009). Ambas as propriedades mecânicas (resistência e deformabilidade) foram determinadas através de um ensaio de compressão uniaxial com medição local das deformações. Todas as amostras que compõem cada uma das bases de dados podem ser agrupadas por formulação em função dos valores da razão Água/Cimento ( $A/C$ ), teor em cimento (massa de cimento/(massa de solo+massa de cimento)) ( $C$ ), tipo de cimento ( $s$ ) e propriedades do solo, isto é, percentagem de areia ( $\%Areia$ ), percentagem de silte ( $\%Silte$ ), percentagem de argila ( $\%Argila$ ) e percentagem de matéria orgânica ( $\%MO$ ). O Quadro 1 resume o número de registos de cada uma das bases de dados utilizadas durante a aprendizagem dos modelos, bem como o respetivo número de formulações. Apesar de o número de registos das bases de dados para o estudo de  $E_{sec50\%}$ ,  $E_{tg50\%}$  e  $E_{max}$  ser relativamente pequeno (apenas 48 amostras), deve-se realçar que a preparação de cada uma das amostras consome bastantes recursos, nomeadamente monetários, tempo, materiais (p.e. cimento) e trabalho laboratorial. A escolha das variáveis dos modelos foi efetuada tendo em conta a informação disponível bem como o conhecimento existente relativo a amostras de solo-cimento. Assim, adicionalmente às sete variáveis previamente enumeradas, a relação entre a porosidade da mistura e o teor volumétrico de cimento ( $n/(C_{lv})^d$ ) e a idade da mistura ( $t$ , dias) foram também consideradas como variáveis do modelo. Nos Quadros 2 e 3 apresentam-se as principais estatísticas de todas as bases de dados utilizadas durante a fase de treino. Todos os solos utilizados durante a preparação das amostras foram alvo de ensaios laboratoriais no sentido de se proceder à sua caracterização. Assim, foram recolhidas amostras do solo natural e testadas em laboratório. Apesar de todos os solos utilizados serem de natureza argilosa, contêm diferentes percentagens de areia, silte, argila e matéria orgânica. No Quadro 4 é apresentada uma descrição detalhada de todos os solos utilizados. Neste quadro, a primeira coluna ("Local") representa o local onde a amostra de solo foi recolhida, enquanto a terceira ("Frequência") indica o número de amostras preparadas para cada tipo de solo. Por último, importa ainda a referir que todas as amostras foram preparadas apenas com três tipos de cimento: CEM I 42.5R ( $s=0.20$ ), CEM II 42.5R ( $s=0.20$ ) e CEM IV/A 35.5R ( $s=0.25$ ).

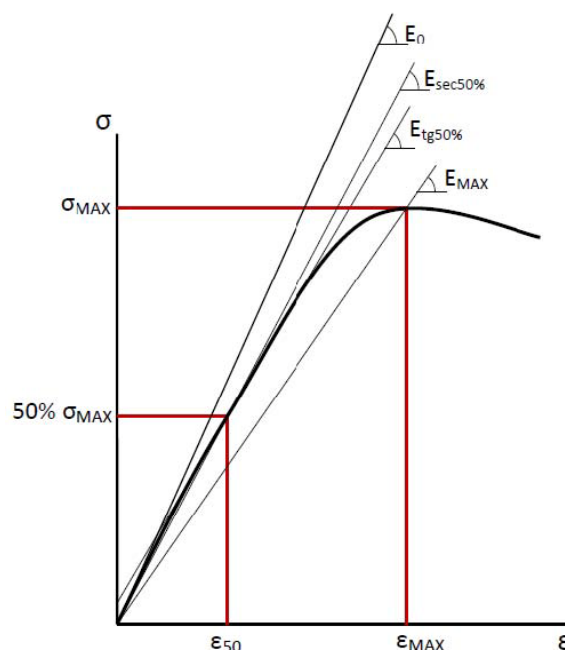


Figura 4 – Ilustração dos Diferentes Módulos de Deformabilidade que Podem ser Definidos num Ensaio de Compressão não Confinado com Medição Local das Deformações

Quadro 1- Número de Registos e Formulações de Cada uma das Bases de Dados Utilizadas

	$Q_u$	$E_0$	$E_{sec50\%}$	$E_{tg50\%}$	$E_{max}$
Número de registos	175	188	48	48	48
Número de formulações	35	9	8	8	8

Quadro 2- Resumo das Principais Estatísticas dos Parâmetros Numéricos das bases de dados da  $Q_u$  e do  $E_0$

Variável	Máximo	Mínimo	Média	Desvio Padrão
$t$				
$Q_u$	56.00	3.00	21.60	19.24
$E_0$	56.00	3.00	20.22	10.07
$C$				
$Q_u$	0.74	0.24	0.47	0.15
$E_0$	0.65	0.24	0.45	0.11
$n/(C_{iv})^d$				
$Q_u$	74.26	48.83	62.59	7.26
$E_0$	75.04	51.21	64.80	7.80
$A/C$				
$Q_u$	1.12	0.68	0.88	0.16
$E_0$	1.11	0.69	0.98	0.12
$s$				
$Q_u$	0.25	0.20	0.21	0.02
$\%Areia$				
$Q_u$	39.00	0.00	13.57	11.53
$E_0$	39.00	0.00	13.44	12.82
$\%Silte$				
$Q_u$	57.00	33.00	50.49	5.49
$E_0$	57.00	33.00	50.57	7.48
$\%Argila$				
$Q_u$	45.00	22.50	35.89	7.74
$E_0$	45.00	22.5	35.85	7.48
$\%MO$				
$Q_u$	8.30	0.40	2.71	1.81
$E_0$	8.30	0.40	3.51	2.28
$Q_u$	13.19	0.76	5.20	2.73
$E_0$	7.89	0.25	2.36	1.32

Quadro 3- Resumo das Principais Estatísticas dos Parâmetros Numéricos das bases de dados do  $E_{sec50\%}$ ,  $E_{tg50\%}$  e  $E_{max}$

Variável	Máximo	Mínimo	Média	Desvio Padrão
$t$	84.00	28.00	64.75	19.29
$C$	0.65	0.24	0.45	0.12
$n/(C_{iv})^d$	75.04	51.21	64.36	7.82
$A/C$	1.11	0.69	0.98	0.12
$\%Areia$	39.00	0.00	14.40	13.67
$\%Silte$	57.00	33.00	49.90	8.32
$\%Argila$	45.00	22.50	35.52	7.40
$\%MO$	8.30	0.40	3.70	2.45
$E_{sec50\%}$	5.67	1.50	3.17	1.11
$E_{tg50\%}$	4.90	1.30	2.76	0.93
$E_{max}$	7.00	1.50	3.44	1.30

Quadro 4- Tipos de Solo Utilizados na Preparação das Amostras Utilizadas na Fase de Treino dos Modelos

Local	Tipo de solo	Frequência					%Areia	%Silte	%Argila	%MO
		$Q_u$	$E_0$	$E_{sec50\%}$	$E_{tg50\%}$	$E_{max}$				
A	Argila Magra (CL)	10	28	9	9	9	39.0	33.0	27.0	8.3
B	Argila Magra Orgânica (OL)	5	18	6	6	6	6.0	57.0	37.0	1.8
C	Argila Gorda (CH)	85	93	22	22	22	7.0	53.0	40.0	3.2
D	Argila Siltosa (CL-ML)	20	27	6	6	6	25.0	52.5	22.5	0.4
E	Argila Magra (CL)	15	22	5	5	5	0.0	55.0	45.0	3.9
F	Argila Siltosa (CL-ML)	20	-	-	-	-	32.5	43.5	24.0	1.2
G	Argila Magra (CL)	20	-	-	-	-	10.5	48.5	41.0	1.0

## 5 - ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nesta secção serão apresentados, analisados e discutidos os resultados obtidos do estudo de previsão das propriedades mecânicas (resistência e deformabilidade) de FLJG por aplicação de técnicas de DM, particularmente do algoritmo MVS.

No Quadro 5 são apresentados os valores das métricas MAD, RMSE e  $R^2$  (valor médio e intervalo de confiança 95%) dos modelos treinados com algoritmo MVS, evidenciando-se para todos eles uma excelente capacidade previsional. Os valores médios (e respetivo intervalo de confiança 95% de acordo com uma distribuição t-student) dos hiperparâmetros de todos os modelos ajustados, bem como o tempo de cálculo são apresentados no Quadro 6. O elevado performance obtido pode ser visualmente observado na Figura 5, onde se representa a relação entre os valores experimentais do  $E_0$  e os previstos pelo respetivo modelo MVS. Como se pode observar, há uma grande proximidade de todos os pontos em relação à linha diagonal que representa a previsão ideal. A relação observada nesta figura para os modelos MVS de previsão da  $Q_u$  e  $E_{tg50\%}$  (ver Tinoco et al. 2010), bem como do  $E_{sec50\%}$  e  $E_{max}$  são em tudo semelhante. Na Figura 5 estão também representadas a curvas REC dos modelos MVS de previsão do  $E_0$ ,  $E_{sec50\%}$ ,  $E_{tg50\%}$  e  $E_{max}$  (lidas nos eixos vertical à direita e horizontal superior), onde se pode observar um rápido aumento da precisão do modelo em função do erro tolerado. Por exemplo, o modelo MVS de previsão do  $E_0$  consegue estimar cerca de 90% dos registos com um desvio absoluto inferior a 0.4GPa.

Em suma, pode-se concluir que ambas as propriedades mecânicas (resistência e deformabilidade) de FLJG podem ser estimadas com elevada precisão pelo respetivos modelos MVS ajustados. Estes modelos permitem assim, conhecendo apenas alguns parâmetros da mistura e as propriedades do solo, prever, ainda em fase de projeto e sem a realização de amostras experimentais, as respetivas propriedades mecânicas ao longo do tempo.

Quadro 5- Medidas de Desempenho para Todos os Modelos Ajustados (média e intervalo de confiança 95%)

	$Q_u$	$E_0$	$E_{sec50\%}$	$E_{tg50\%}$	$E_{max}$
MAD	0.55±0.00	0.17±0.00	0.15±0.01	0.15±0.00	0.18±0.00
RMSE	0.73±0.00	0.25±0.01	0.21±0.03	0.20±0.00	0.31±0.01
$R^2$	0.93±0.00	0.96±0.00	0.96±0.01	0.95±0.00	0.94±0.00

Quadro 6- Hiperparâmetros e Tempo de Cálculo para todos os Modelos Ajustados (média e intervalo de confiança 95%)

	$Q_u$	$E_0$	$E_{sec50\%}$	$E_{tg50\%}$	$E_{max}$
gamma	0.55±0.00	0.17±0.00	0.15±0.01	0.15±0.00	0.18±0.00
$\epsilon$	0.73±0.00	0.25±0.01	0.21±0.03	0.20±0.00	0.31±0.01
Tempo (s)	0.93±0.00	0.96±0.00	0.96±0.01	0.95±0.00	0.94±0.00

Na análise de um modelo, adicionalmente à sua capacidade previsional é também muito importante a sua interpretação. No caso de modelos obtidos por aplicação de técnicas de DM, nomeadamente pelo algoritmo MVS, tal tarefa apresenta uma dificuldade acrescida devido à elevada complexidade das expressões matemáticas envolvidas. No presente estudo, a interpretabilidade dos modelos foi efetuada através da realização de uma análise de sensibilidade detalhada. Tal análise permitiu, por um lado identificar a importância relativa de cada uma das variáveis envolvidas no modelo, e por outro perceber o efeito que cada uma tem no estudo da respetiva propriedade. Além disso, permitiu ainda identificar qual o nível de interação entre variáveis. Os resultados a seguir apresentados procuram explicitar os modelos ajustados contribuindo para uma melhor compreensão do comportamento mecânico de FLJG.

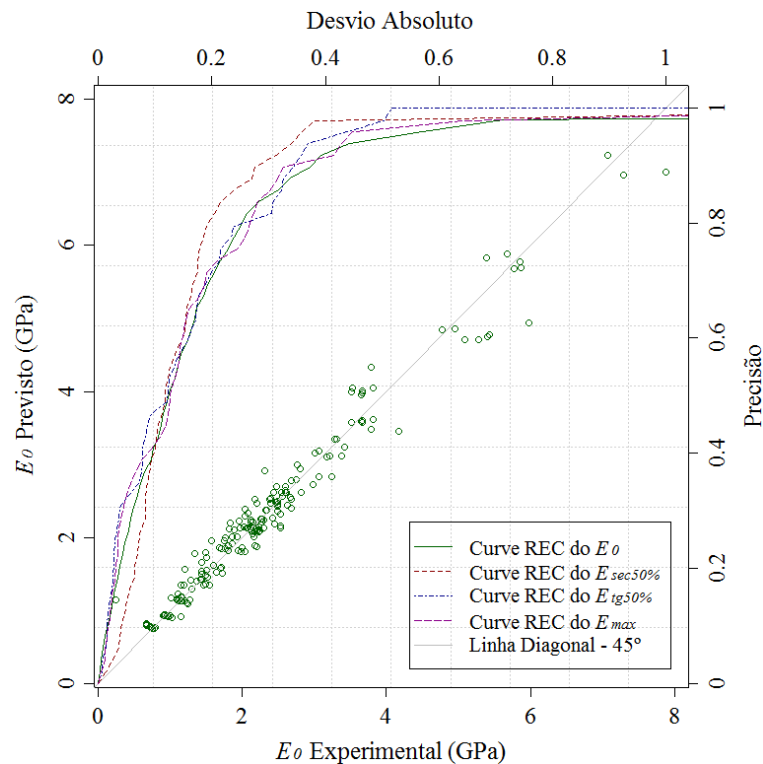


Figura 5 – Relação entre os Valores Experimentais do  $E_0$  versus Previstos pelo modelo MVS e Curvas REC dos modelos de Previsão da Deformabilidade de FLJG. Estas Curvas Deverão ser Lidas no Eixo Superior Horizontal e Vertical à Direita

Após a realização de uma análise de sensibilidade unidimensional, onde se faz alterar apenas uma variável de cada vez, mantendo as restantes nos respetivos valores médios, quantificou-se a importância relativa de cada atributo. Na Figura 6 compara-se a importância relativa de cada variável para cada um dos modelos treinados. Analisando esta figura, é possível observar que a relação  $n/(C_{IV})^d$  é a variável que apresenta maior impacto no estudo das propriedades mecânicas de FLJG (16% no estudo da resistência e cerca de 30% no estudo da deformabilidade). No caso do estudo da resistência as variáveis  $t$ ,  $C$  e  $n/(C_{IV})^d$  são as três variáveis chave. Já no estudo da deformabilidade, as propriedades do solo, nomeadamente a respetiva percentagem em argila, seguem a relação  $n/(C_{IV})^d$  nas variáveis com maior impacto. Uma observação também interessante, no caso do estudo da deformabilidade, prende-se com o facto de o fator tempo apresentar um efeito relativamente pequeno, exceto no estudo do  $E_0$ . Mais adiante procurar-se-á justificar esta observação. Da análise da Figura 6 importa ainda a referir o menor efeito das propriedades do solo no estudo da resistência quando comparado com o estudo da deformabilidade.

No sentido de melhorar a interpretabilidade dos modelos, foi quantificado o efeito de cada atributo no estudo das propriedades mecânicas de FLJG. Assim, e uma vez mais através da realização de uma análise de sensibilidade unidimensional foram construídas as curvas VEC para as variáveis mais relevantes no estudo da resistência e deformabilidade de FLJG. Na Figura 7 é ilustrada a curva REC, e respetivo histograma (frequências lidas no eixo vertical à direita), relativa à variável  $t$  evidenciando um efeito não linear. Tal como expectável, tendo em conta o conhecimento empírico relativo a misturas solo-cimento, o fator  $t$  apresenta um efeito positivo na previsão da  $Q_u$  e a forma convexa da curva REC mostra que o seu efeito é mais preponderante para idades iniciais e tende a estabilizar para idades avançadas (acima dos 45 dias de cura).

Na Figura 8 ilustram-se as curvas VEC da variável mais relevante no estudo da deformabilidade de FLJG, ou seja,  $(n/(C_{IV})^d)$ , comparando-se o respetivo formato para cada um dos diferentes módulos. As quatro curvas VEC apresentam um aspeto semelhante evidenciando um efeito negativo na previsão da deformabilidade de FLJG. Isto significa que amostras com elevada porosidade ou baixos teores em cimento apresentam menor rigidez, particularmente para valores de  $n/(C_{IV})^d$  superiores a 60.



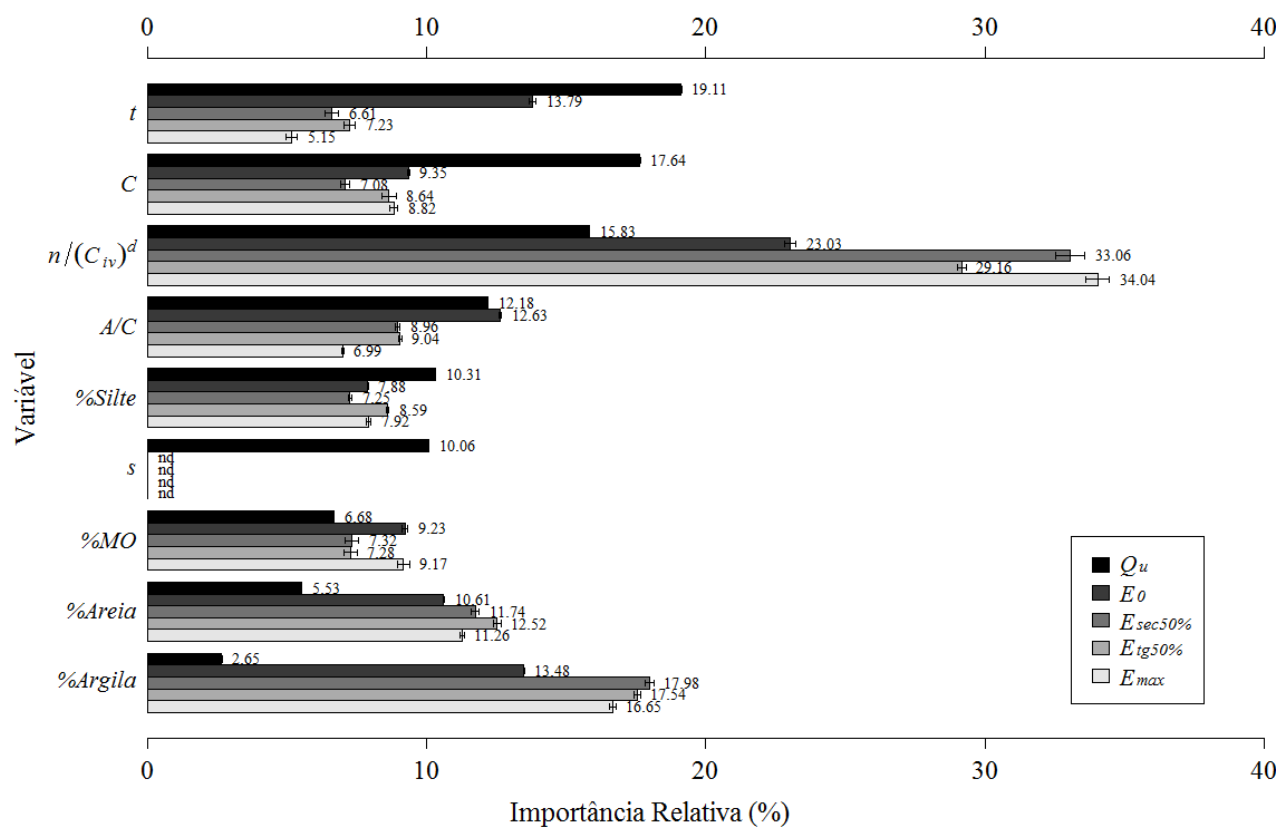


Figura 6 – Comparação da Importância Relativa de Cada uma das Variáveis no Estudo das Propriedades Mecânicas de FLJG

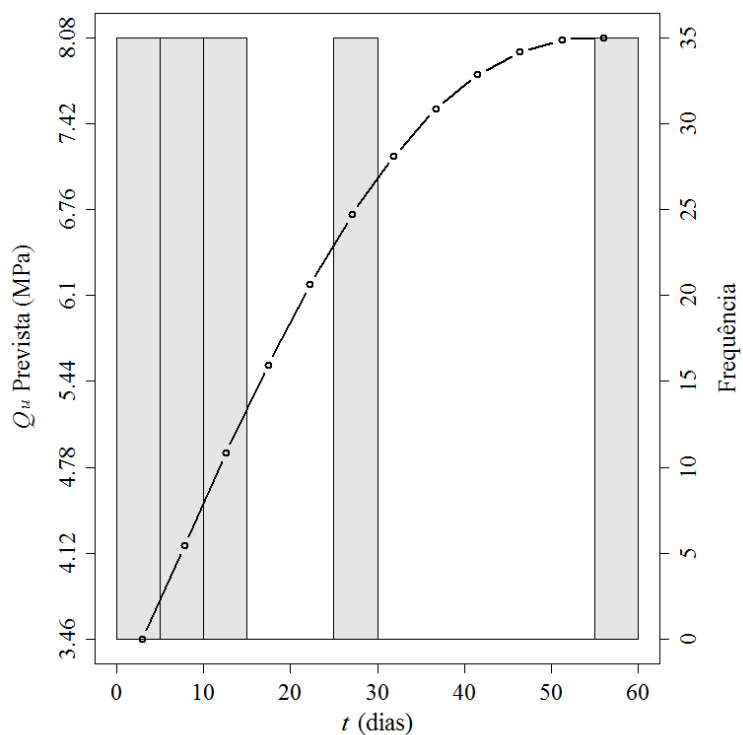


Figura 7 – Histograma (barras) e Respetiva Curva VEC da Variável  $t$  de Acordo com o Modelo MVS de Previsão da  $Q_u$  de FLJG, Quantificadas por uma Análise de Sensibilidade Unidimensional

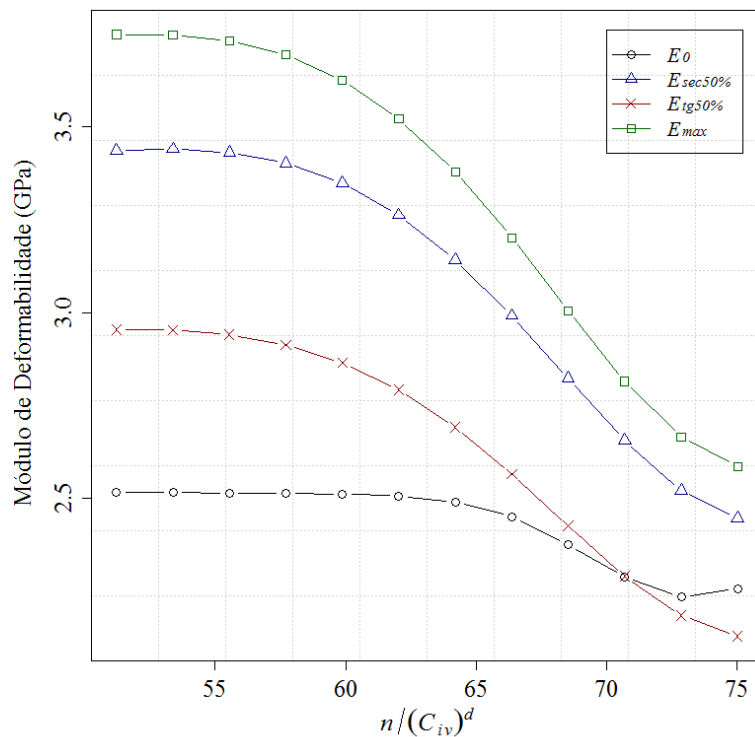


Figura 8 – Curvas VEC da variável  $n/(C_{iv})^d$  Segundo os Modelos MVS de Previsão da Deformabilidade de FLJG, Quantificadas por uma Análise de Sensibilidade Unidimensional

Todos os resultados previamente apresentados (importância relativa e curvas VEC) foram obtidos através da realização de análises de sensibilidade unidimensionais, isto é, fazendo variar apenas uma variável de cada vez mantendo as restantes nos respectivos valores médios. Contudo, este tipo de condições raramente acontece em problemas reais. Assim, visando uma análise mais realista e detalhada, nos parágrafos seguintes são apresentadas e discutidas algumas observações resultantes da realização de uma análise de sensibilidade bidimensional, isto é, fazendo alterar duas variáveis em simultâneo.

Na Figura 9 é apresentado o nível de interação de todas as variáveis com a variável mais relevante no estudo da  $Q_u$ , isto é, com  $t$ . Da sua análise é possível verificar que  $A/C$  é a variável com maior interação com  $t$ , evidenciando uma importância relativa de aproximadamente 14%. Esta observação mostra que a  $A/C$  apesar de ser apenas a quarta variável com maior impacto na previsão da  $Q_u$  de FLJG (ver Figura 6) deverá ser levada em consideração no estudo do comportamento mecânico deste material. Esta forte interação pode de certa forma ser explicada se se levar em consideração o facto de o aumento de resistência de misturas solo-cimento estar relacionado com a diminuição de água livre na mistura (processo de hidratação do cimento). Is significa que no caso de amostras preparadas com  $A/C$  mais elevadas será necessário esperar mais tempo para obter a mesma resistência do que em amostras preparados com  $A/C$  inferiores. Por outro lado, em amostras preparadas com  $A/C$  mais baixas o processo de endurecimento é mais rápido.

Representando o efeito da interação entre  $t$  e  $A/C$  na previsão da  $Q_u$ , obtém-se a superfície VEC ilustrada na Figura 10. Esta superfície mostra precisamente o elevado efeito da interação entre estas duas variáveis, evidenciado pelo elevado intervalo de variação dos valores da  $Q_u$  para as diferentes combinações entre  $t$  e  $A/C$  (desde 2MPa até 9MPa). Além disso, é possível observar que misturas com elevada  $A/C$  tendem a estabilizar para idades mais recentes.

No estudo da deformabilidade de FLJG foi realizada uma análise de sensibilidade bidimensional no sentido de medir o nível de interação de todas as variáveis com a relação  $n/(C_{iv})^d$  para cada um dos quatro diferentes módulos estudados. No Quadro 7 é apresentado o nível de interação de todas as variáveis com  $n/(C_{iv})^d$ , podendo-se observar que a %Argila é aquela cuja interação maior impacto tem no estudo do respetivo módulo. A variável tempo apresenta também uma forte interação com  $n/(C_{iv})^d$ , sendo mesmo superior à da %Argila no caso do  $E_0$ . Esta forte interação da variável  $t$  com  $n/(C_{iv})^d$  (variável chave no estudo da deformabilidade de FLJG) permite de certa forma justificar a menor importância relativa da variável  $t$  quantificada através da realização de uma análise de sensibilidade unidimensional (ver Figura 6). Na Figura 11 representa-se o efeito da interação entre  $n/(C_{iv})^d$  e a %Argila no estudo do  $E_{tg50\%}$ . Da sua análise verifica-se que os maiores valores de  $E_{tg50\%}$  são obtidos em amostras com reduzida  $n/(C_{iv})^d$  e preparados com solos onde o teor em argila é também reduzido. Verifica-se ainda que  $E_{tg50\%}$  é mais

sensível a variações de  $n/(C_{iv})^d$  do que ao teor em argila do solo. A mesma representação para os restantes módulos de deformabilidade é em tudo semelhante.

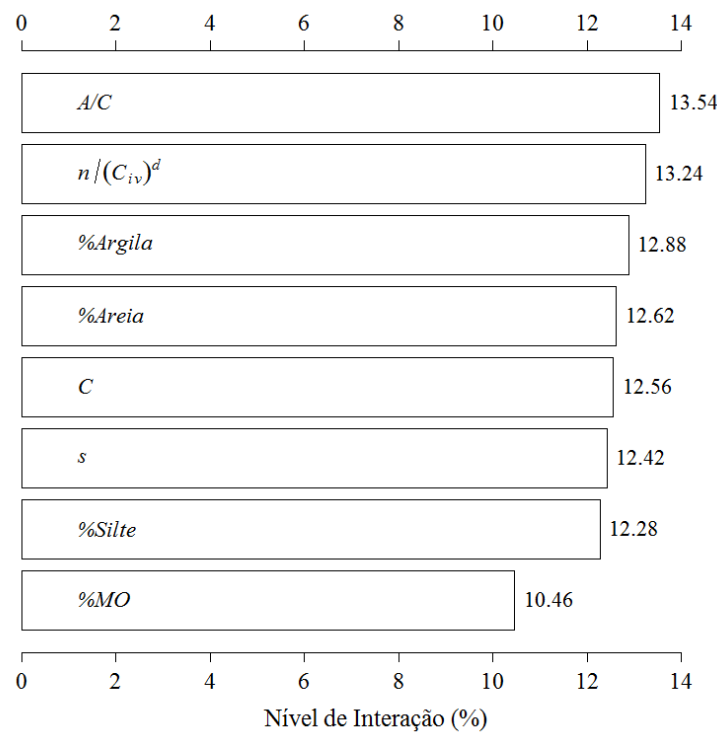


Figura 9 – Nível de Interação de Todas as Variáveis com  $t$ , de acordo com o Modelo MVS de Previsão da  $Q_u$  de FLJG, após a Aplicação de uma Análise de Sensibilidade Bidimensional

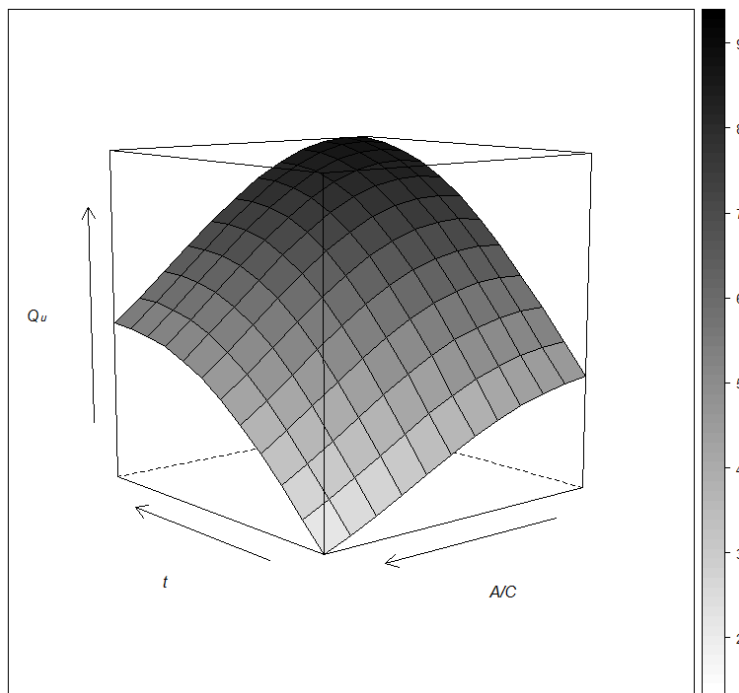


Figura 10 – Superfície VEC da Interação entre  $t$  e  $A/C$  do estudo da  $Q_u$  de FLJG, Quantificadas por uma Análise de Sensibilidade Bidimensional

Quadro 7- Nível de Interação de Todas as Variáveis com  $n/(C_{iv})^d$ , de acordo com os Modelo MVS de Previsão de  $E_0$ ,  $E_{sec50\%}$ ,  $E_{tg50\%}$  e  $E_{max}$  de FLJG, após a Aplicação de uma Análise de Sensibilidade Bidimensional

	<i>t</i>	<i>C</i>	<i>A/C</i>	%Areia	%Silte	%Argila	%MO
<i>n/(C<sub>iv</sub>)<sup>d</sup></i>	<i>E<sub>0</sub></i>	21.22	10.60	14.48	15.86	10.45	17.94
	<i>E<sub>sec50%</sub></i>	15.57	13.49	13.18	14.92	12.44	16.31
	<i>E<sub>tg50%</sub></i>	15.34	13.20	13.39	15.00	12.46	16.48
	<i>E<sub>max</sub></i>	15.80	13.77	12.89	14.90	12.42	16.16

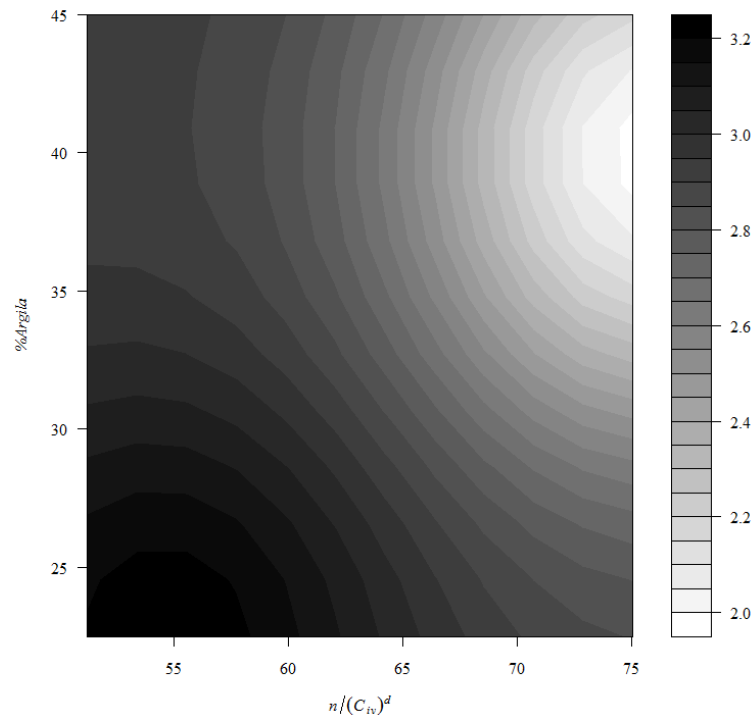


Figura 11 – Contorno VEC da Interação entre  $n/(C_{iv})^d$  e  $\%Argila$  do estudo do  $E_{tg50\%}$  de FLJG, Quantificadas por uma Análise de Sensibilidade Bidimensional

## 6 - CONCLUSÕES

No contexto da tecnologia de *jet grouting* (JG) a principal desvantagem está relacionada com a ausência de modelos racionais para o dimensionamento das propriedades mecânicas da mistura obtida, nomeadamente a resistência à compressão uniaxial ( $Q_u$ ) e deformabilidade. Atualmente, esta tarefa é essencialmente realizada considerando os resultados de ensaios laboratoriais e modelos empíricos. No entanto, nas fases iniciais de projeto, principalmente em obras geotécnicas de pequena envergadura, a escolha dos valores a adotar para cada um dos parâmetros é baseada em informação muito escassa. Assim, o uso de dados de projetos já realizados pode ser visto como uma alternativa para ultrapassar este problema. A aplicação de técnicas de *data mining* (DM) a dados bem organizados coletados em grandes obras geotécnicas pode fornecer uma importante base para desenvolver modelos que poderão ser muito úteis em futuros projetos.

No presente artigo, foram propostos modelos racionais de previsão da  $Q_u$  e dos diferentes módulos de deformabilidade, isto é, módulo de elasticidade inicial ( $E_0$ ), módulo de deformabilidade secante a 50% da tensão de ruptura ( $E_{sec50\%}$ ), módulo de deformabilidade tangente a 50% da tensão de ruptura ( $E_{tg50\%}$ ) e módulo de deformabilidade máximo ( $E_{max}$ ), de formulações laboratoriais de *jet grouting* (FLJG) ao longo do tempo por aplicação de técnicas de DM. Os modelos propostos são capazes de estimar com elevado grau de precisão a  $Q_u$  e o  $E_0$  entre os 3 a 56 dias, bem como o  $E_{sec50\%}$ ,  $E_{tg50\%}$  e  $E_{max}$  entre 28 e 84 dias de FLJG, preparadas com solos argilosos contendo diferentes percentagens de areia, silte, argila e matéria orgânica. Os resultados obtidos mostraram que o algoritmo máquinas de vetores de suporte (MVS) é capaz de aprender as complexas relações entre as propriedades mecânicas de FLJG e as diferentes variáveis intervenientes. A realização de uma análise de sensibilidade unidimensional sobre os diferentes modelos treinados permitiu estabelecer algumas importantes observações. A relação entre a porosidade da mistura e o teor volumétrico de cimento ( $n/(C_{iv})^d$ ) foi identificada como a variável chave no estudo das propriedades mecânicas de FLJG. No caso do estudo da  $Q_u$ , a variável tempo e o teor em cimento da mistura são também variáveis muito importantes. Observou-se ainda que o efeito das variáveis mais relevantes no estudo da  $Q_u$  é predominantemente não linear. No caso da variável tempo, o seu efeito segue uma lei exponencial, corroborando o conhecimento empírico relativo a misturas solo-cimento. Quando o objetivo passa por estimar a deformabilidade de FLJG, as propriedades do solo natural, principalmente o seu teor em argila, deverá ser considerado como uma variável do modelo. Também aqui se observou um comportamento não linear das principais variáveis intervenientes, que no caso da relação  $n/(C_{iv})^d$ , e de acordo com o que seria espetável, o seu aumento conduz a uma diminuição da rigidez das misturas. A realização de uma análise de sensibilidade bidimensional sobre os modelos de previsão da deformabilidade de FLJG mostrou que existe uma forte interação entre a idade da mistura e a relação ( $n/(C_{iv})^d$ ), justificando o menor impacto da primeira no estudo da deformabilidade aquando da realização de uma análise de sensibilidade unidimensional.



Face ao excelente desempenho dos modelos treinados, não apenas em termos de capacidade previsional mas também em termos de coerência com o atual conhecimento empírico relativo ao comportamento mecânico de FLJG, pode-se concluir que tais modelos podem ser utilizados para descrever com elevado grau de fiabilidade o conhecimento empírico relativo a este tipo de material e servir de apoio ao dimensionamento das respetivas propriedades mecânicas (resistência e deformabilidade). Todo o conhecimento obtido pode ser visto como um importante avanço no domínio da engenharia geotécnica, na medida em que contribuirá para o aumento do conhecimento sobre misturas de JG, o que permitirá uma redução significativa do número de formulações laboratoriais necessárias, melhorando a eficiência técnica e económica do tratamento.

Em trabalhos futuros, estas ferramentas de DM serão aplicadas no desenvolvimento de modelos racionais de previsão das propriedades mecânicas (resistência e deformabilidade) de amostras extraídas diretamente de colunas de JG. Serão ainda aplicadas as mesmas metodologias na previsão do diâmetro de colunas de JG, onde é expectável um bom desempenho tendo em conta algumas experiências preliminares. Face aos resultados obtidos, a extensão das metodologias aplicadas no estudo de outros tipos de materiais, nomeadamente àqueles resultantes de tratamentos por aplicação da tecnologia de Deep mixing, poderá também ser equacionada em linhas de investigação futuras.

## AGRADECIMENTOS

Os autores gostariam de agradecer à Fundação Para a Ciência e a Tecnologia (FCT) pelo apoio financeiro no âmbito do projeto PEst-OE/ECI/UI4047/2011 e pela bolsa de doutoramento SFRH/BD/45781/2008. Agradeces-e também à empresa Tecnasol-FGE por todo o interesse e apoio prestado.

## REFERÊNCIAS

- Bi, J. e Bennett, K. (2003). Regression Error Characteristic Curves (CD-ROM). *Proceeding of Twentieth International Conference on Machine Learning*. AAAI Press, Washington, DC, pp.40-50.
- Chen, Y. e Councill, I. (2003). An Introduction to Support Vector Machines: A Review. *AI Magazine*. AAAI Press, Vol.24(2), pp.105-107.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. e Wirth, R. (2000). *CRISPDM1.0: Step-by-Step Data Mining Guide*. CRISP-DM consortium, 78 p.
- Cherkassky, V. e Ma, Y. (2004). Practical Selection of SVM Parameters and Noise Estimation for SVM Regression. *Neural Networks*. Elsevier, Vol.17(1), pp.113-126.
- Cortes, C. e Vapnik, V. (1995). Support Vector Networks. *Machine Learning*. Springer, Vol.20(3), pp.273-297.
- Cortez, P. (2010). Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool. *Advances in Data Mining – Applications and Theoretical Aspects 10th Industrial Conference on Data Mining (ICDM 2010), Lecture Notes in Artificial Intelligence 6171*. Springer, Berlin, Alemanha, pp. 572-583.
- Cortez, P. e Embrechts, M. (2011). Opening Black Box Data Mining Models Using Sensitivity Analysis. *Proceedings of the 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2011)*. IEEE, Paris, França, pp.341-348.
- Croce, P. e Flora, A. (2001). Analysis of single-fluid jet grouting. *Géotechnique*, Vol.51, pp.905-906.
- Fayyad, U., Piatetsky-Shapiro, G. e Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*. ACM, Vol.39(11), pp. 27-34.
- Frohlich, H. e Zell, A. (2005). Efficient Parameter Selection for Support Vector Machines in Classification and Regression via Model-Based Global Optimization. *Neural Networks*. IEEE, Vol.3, pp.1431-1436.
- Gazaway, H. e Jasperse, B. (1992). Jet grouting in Contaminated Soils. *Proceedings of Grouting, Soil Improvement and Geosynthetics (GSP 30)*. ASCE, New Orleans, LA, USA, pp.206-214.
- Gomes Correia, A., Valente, T., Tinoco, J., Falcão, J., Barata, J., Cebola, D., e Coelho, S. (2009). Evaluation of mechanical properties of jet grouting columns using different test methods. *Proceedings of 17th International Conference on Soil Mechanics and Geotechnical Engineering*. IOS Press, Alexandria, Egito, pp.2179-2171.
- Goh, A. e Goh, S. (2007). Support Vector Machines: Their use in Geotechnical Engineering as Illustrated using Seismic Liquefaction Data. *Computers and Geotechnics*. Elsevier, Vol.34(5), pp.410-421.
- Hastie, T., Tibshirani, R. e Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed, Springer-Verlag, 745 p.

- Huang, C., Lee, Y., Lin, D. e Huang, S. (2007). Model Selection for Support Vector Machines via Uniform Design. *Computational Statistics & Data Analysis*. Elsevier, Vol.52(1), pp.335-346.
- Kirsh, F. e Wolfgang, S. (2001). Ground Improvement and its Numerical Analysis. *Proceedings of 15th International Conference on Soil Mechanics and Geotechnical Engineering*. AA Balkema Publishers, Atenas, Grécia, Vol.3, pp.1775-1778.
- Lee, F., Lee, Y., Chew, S. e Yong, K. (2005). Strength and modulus of marine clay-cement mixes. *Journal of geotechnical and geoenvironmental engineering*. ASCE, Vol.131(2), pp.178-186.
- Liu, S., Zhang, D., Liu, Z. e Deng, Y. (2008). Assessment of Unconfined Compressive Strength of Cement Stabilized Marine Clay. *Marine Georesources & Geotechnology*. Taylor & Francis, Vol.26(1), pp.19-35.
- Miranda, T., Gomes Correia, A., Santos, M., Sousa, L. e Cortez, P. (2011). New Models for Strength and Deformability Parameter Calculation in Rock Masses Using Data-Mining Techniques. *International Journal of Geomechanics*. ASCE, Vol.11, pp.44-58.
- Momma, M. e Bennett, K. (2011). A Pattern Search Method for Model Selection of Support Vector Regression. *Proceedings of the SIAM International Conference on Data Mining*. Philadelphia: SIAM, Arizona, USA, pp.261-274.
- Narendra, B., Sivapullaiah, P., Suresh, S. e Omkar, S. (2006). Prediction of unconfined compressive strength of soft grounds using computational intelligence techniques: A comparative study. *Computers and Geotechnics*. Elsevier, Vol.33(3), pp.196-208.
- Nikbakhtan, B. e Ahangari, K. (2010a). Field study of the influence of various jet grouting parameters on soilcrete unconfined compressive strength and its diameter. *International Journal of Rock Mechanics and Mining Sciences*. Elsevier, Vol.47, pp.685-689.
- Nikbakhtan, B., Ahangari, K. e Rahmani, N. (2010b). Estimation of Jet Grouting Parameters in Shahriar Dam, Iran. *Mining Science and Technology*. Elsevier, Vol.20(3), pp.472-477.
- R Development Core Team (2009). *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>, acedido em 07/12/2011.
- Shibazaki, M. (2003). State of practice of jet grouting. *Proceedings of Grouting and Ground Treatment: 3rd International Specialty Conference on Grouting and Ground Treatment*. ASCE, New Orleans, Louisiana, USA, pp.198-217.
- Smola, A. e others (1996). *Regression Estimation with Support Vector Learning Machines*. Master's Thesis, Technische Universit at Munchen, 78 p.
- Smola, A. e Schölkopf, B. (2004). A Tutorial on Support Vector Regression. *Statistics and Computing*. Springer, Vol.14(3), pp.199-222.
- Tinoco, J., Gomes Correia, A. e Cortez, P. (2010). Aplicação de Técnicas de Data Mining na Previsão do Comportamento Mecânico de Colunas de Jet Grouting. *Atas do XII Congresso Nacional de Geotecnia*. A. Gomes Correia et al. (Eds.). Guimarães, Portugal, pp. 2167–2176.
- Tinoco, J., Gomes Correia, A. e Cortez, P. (2011a). Using Data Mining Techniques to Predict Deformability Properties of Jet Grouting Laboratory Formulations over Time. *Proceeding of 15th Portuguese Conference on Artificial Intelligence*. L. Antunes and H.S. Pinto (Eds.). Springer-Verlag, Guimarães, Portugal, pp. 491–505.
- Tinoco, J., Gomes Correia, A. e Cortez, P. (2011b). Uniaxial Compressive Strength Prediction of Jet Grouting Columns using Support Vector Machines. *Proceedings of The 2011 European Simulation and Modeling Conference*. EUROSIS, Guimarães, Portugal, pp.326-330.
- Vapnik, V. N. (2000). *The nature of statistical learning theory*, 2nd ed., Springer-Verlag, New York, USA, 314 p.
- Xanthakos, P., Abramson, L. e Bruce, D. (1994). *Ground Control and Improvement*. Wiley-Interscience.
- Welsch, J. e Burke, G. (1991). Jet Grouting – Uses For Soil Improvement. *Proceedings of Geotechnical Engineering Congress (GSP No. 27)*. pp.334-345.

## XIII Congresso Nacional de Geotecnia

Pensar e Construir com a Natureza. Uma Visão para a Engenharia

Conferências Especiais | Conhecer. Os Geomateriais | Pensar.

A Concepção e o Projecto | Usufruir. A Manutenção e a Exploração |

Desafiar. Como Lidar com o Risco | Inovar. Perspectivas

na Geotecnia – Ligação entre a Ciência e Tecnologia e Sociedade

3º Encontro de Jovens Geotécnicos

## VI Congresso Luso-Brasileiro de Geotecnia

Risco e Segurança

Taludes Naturais e Ações Climáticas | Geossintéticos: Uso Ambiental |

Melhoria do Terreno: Inovações e Previsões de Comportamento |

Mecânica dos Solos Rijos e das Rochas Brandas

